Genie: Generative Interactive Environments

Year: 2024 | Citations: 328 | Authors: Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi

---

Abstract

We introduce Genie, the first generative interactive environment trained in an unsupervised manner from unlabelled Internet videos. The model can be prompted to generate an endless variety of action-controllable virtual worlds described through text, synthetic images, photographs, and even sketches. At 11B parameters, Genie can be considered a foundation world model. It is comprised of a spatiotemporal video tokenizer, an autoregressive dynamics model, and a simple and scalable latent action model. Genie enables users to act in the generated environments on a frame-by-frame basis despite training without any ground-truth action labels or other domain-specific requirements typically found in the world model literature. Further the resulting learned latent action space facilitates training agents to imitate behaviors from unseen videos, opening the path for training generalist agents of the future.